The role of non-coding RNAs in B-cell lymphoma

Tayari, Masoumeh

CHAPTER 5

eQTL analysis of candidate susceptibility loci in Hodgkin lymphoma

Mina Masoumeh Tayari, Rianne Veenstra, Lydia Visser, Arjan Diepstra, Joost Kluiver and Anke van den Berg.

Department of Pathology and Medical Biology, University of Groningen, University Medical Center Groningen, Groningen, the Netherlands

Manuscript in preparation

ABSTRACT

Genome-wide association studies have reported multiple Hodgkin lymphoma (HL) susceptibility loci. However, the susceptibility mechanisms of the alleles are unknown. Here we evaluate possible expression quantitative trait locus (eQTL) effects of four susceptibility loci harboring besides a protein-coding gene also one or multiple long non-coding (lnc)RNAs, i.e. 2p16.1 (*REL, TCONS_00002901* and *LINC01185*), 5q31 (*IL13* and *TH2LCRR*), 8q24.21 (*MYC, PVT1, CASC11* and *LINC00824*) and 10p14 (*GATA3* and *GATA3-AS1*).

Genotype-expression correlations were studied in lymphoblastoid cell lines (LCLs) generated from PBMCs of 98 healthy controls and 72 post-therapy HL patients. This revealed a significant relationship between GATA3 expression levels and the three associated 10p14 SNP genotypes in LCLs of post-therapy cHL patients, but not of healthy controls. For all three SNPs lower GATA3 expression levels were associated with the risk allele. No associations were observed for the GATA3-AS1 lncRNA. No significant relationships were observed between expression levels and SNP genotypes of the protein-coding and/or lncRNA genes at the 2p16.1, 5q31 and 8q24.21 regions in LCLs of healthy controls or post-therapy cHL patients. Comparison of the expression levels of the protein-coding and non-coding transcripts per region revealed a positive correlation between the expression levels of REL and TCONS_00002901 with r^2=0.74 in the control LCLs and r^2=0.25 in post-therapy cHL patient LCLs. Similarly, GATA3 and GATA3-AS1 expression levels also showed a positive correlation that was stronger (r^2=0.71) in control LCLs as compared to post-therapy cHL patient LCLs (r^2=0.29). No correlation was observed between the expression levels of MYC and PVT1, which was the only 8q24.21 lncRNA that was expressed in the LCLs.

Thus, we showed a cis-eQTL effect for the alleles of three 10p14 SNPs with GATA3 expression in LCLs of post-therapy cHL patients. In addition, we showed a positive correlation between expression levels of REL and GATA3 with the nearby lncRNA gene levels.

INTRODUCTION

Classical Hodgkin lymphoma (cHL) is regarded as a complex disease caused by a combination of both environmental and genetic factors. CHL has a strong genetic component with an increased risk of cHL in first and second-degree relatives of cHL patients [1]. Moreover, monozygotic twins of patients with cHL have a greatly increased risk compared to dizygotic twins or other siblings of patients with cHL [2].

To improve our understanding of genetic factors contributing to cHL susceptibility, we and others have performed genome-wide association studies (GWAS) [3-7]. Most of the associated SNPs were located in the human leukocyte antigen (HLA) region on chromosome six (6p21.32 to 6p22.1), which emphasizes the importance of the HLA region in disease etiology. Non-HLA susceptible associations were found at 10 loci, including 2p16.1 [4, 6], 3p24 [6, 7], 3q32 [7], 5q15 [3], 5q31 [3, 7], 6q23 [6, 7], 7p21,3 [7], 8q24.21 [4, 6, 7], 10p14 [4, 6, 7] and 19p13 [7]. These non-HLA associations implicate important roles of genes mapping to these loci in cHL susceptibility. For most of the protein-coding genes at these loci, a role in cHL pathogenesis has already been shown or postulated. In general, SNP alleles of susceptibility loci can affect gene function and / or expression via multiple mechanisms. For the non-HLA region at 19p13, we previously showed an expression quantitative trait loci (eQTL) effect of the associated SNP alleles with TCF3 expression levels in control and post-therapy cHL patient-derived lymphoblastoid cell lines (LCLs) [7]. The IL-13 missense SNP at 5q31 results in a change of arginine (R110) to glutamine (Q110) [8]. The Q110 IL-13 variant has been reported to be more active than wild type IL-13, based on enhanced STAT6 phosphorylation and CD23 expression in monocytes and induction of IgE switching in B-cells [8]. For most of the SNP alleles, it remains unknown how they can contribute to cHL susceptibility.

More than 90% of disease-associated SNPs map to non-coding regions of the genome [9]. Next to mapping to regulatory regions such as promoters,

enhancers and other regulatory elements in the genome these SNPs also map to long non-coding (lnc)RNA loci [10]. Disease-associated SNPs mapping at protein or lncRNA loci may influence disease susceptibility by modulating expression of the corresponding genes (eQTL effect) or affect functionality of the gene products [10, 11]. Hence, the identification of eQTL effects of disease-associated SNPs can contribute to our understanding of susceptibility mechanisms.

We hypothesized that cHL-associated SNPs act as eQTLs of the nearby protein-coding or lncRNA genes. To test this hypothesis, we performed eQTL analysis for 4 of the 10 non-HLA regions known to be associated with cHL in lymphoblastoid cell lines (LCL) genotyped for the associated SNPs. In addition, we investigated a possible cis-regulatory effect of the lncRNAs by correlating their expression levels to the expression levels of the protein-coding genes [12].

MATERIAL AND METHODS

Selection of SNPs

Genetic variants were selected based on previously published cHL GWAS [3, 4, 6, 7]. We focused on four genomic regions that contained besides a candidate protein-coding gene, also one or more lncRNA genes. Information about selected SNPs is given in Table 1.

Table 1. Overview of the SNPs associated with cHL in the four selected loci.

SNP	Gene locus	Gene candidates	cHL association	MAF* in patients	MAF in controls	Reference
rs1432295	2p16.1	REL / TCONS_00002901 / LINC001185	G-risk Minor allele	48 %	40 %	[4, 6]
rs13034020	2p16.1	REL / TCONS_00002901 / LINC001185	G-risk Minor allele	18 %	16 %	[7]
rs2069757	5q31	IL13 / TH2LCRR	A-risk Minor allele	11 %	7 %	[7]
rs20541	5q31	IL13 / TH2LCRR	A-risk Minor allele	24 %	18 %	[3, 7]
rs2019960	8q24.21	MYC / PVT1 / CASC11 / LINC00824	G-risk Minor allele	29 %	23 %	[4, 6, 7]
rs2608053	8q24.21	MYC / PVT1 / CASC11 / LINC00824	G-risk Major allele	41 %	48 %	[4, 6]
rs444929	10p14	GATA3 / GATA3-AS1	C-risk Minor allele	24 %	21 %	[7]
rs485411	10p14	GATA3 / GATA3-AS1	A-risk Minor allele	31 %	25 %	[4, 6]
rs501764	10p14	GATA3 / GATA3-AS1	C-risk Minor allele	25 %	19 %	[4, 6]

* Minor Allele Frequency

Generation of EBV-transformed LCLs and SNP genotyping

LCLs were generated by infection of PBMCs with the Epstein-Barr virus (EBV) strain B95-8 virus as described earlier [13] from 98 healthy controls and 72 post-therapy cHL patients. For cHL patients PBMC were obtained at least 1 year after completion of all therapies. Cyclosporin A, 5ug/ml (1:50) was used to inhibit T cell-mediated killing of infected B cells. LCLs were cultured in RPMI 1640 with 10% FCS.

Genomic DNA was isolated from the LCLs using the salt / chloroform method following standard laboratory protocols. Concentrations of the DNA samples were measured with a Nanodrop™ 1000 Spectrophotometer and DNA integrity was evaluated on a 1% agarose gel. Genotyping of the LCLs was performed using TaqMan SNP assays (Supplementary Table 1), Taqman Universal PCR Master Mix (no AmpErase UNG), and 8-10ng of DNA in a total volume of 5µl. SNP genotyping was performed in triplicate on a ABI PRISM 7900HT and allelic discrimination was performed using the SDS software (Applied Biosystems, Carlsbad, CA). For the healthy control LCL group genotyping failed for one sample for rs1432295, rs20541, rs2608053, rs485411. For the post-therapy cHL patients LCL group genotyping failed for one sample for rs2069757.

RNA isolation and Quantitative RT-PCR

RNA was isolated from LCLs as described earlier using miRNeasy Mini Kit [14]. RNA concentration was measured with a NanoDrop™ 1000 Spectrophotometer (Thermo Fisher Scientific Inc., Waltham, USA) and integrity was assessed by analysis of the 18S/28S bands on a 1% agarose gel. Expression levels were assessed using RT-qPCR as described previously [15]. Primer sequences and gene expression assay ID numbers (Thermo Fisher Scientific Inc., Waltham, USA) are shown in Table S2. Relative expression levels were calculated using TBP as housekeeping gene and data were expressed as the 2-delta Cp values.

Statistical analysis

Differences in the distribution of gene expression levels by SNP genotype were tested using linear regression and one way ANOVA, separately for the post-therapy cHL patient and healthy control LCL groups.

RESULTS

Expression levels of 2p16.1 genes in SNP genotype stratified LCLs

The associated region at 2p16.1 encompasses 137 kb in two linkage disequilibrium (LD) blocks. The two associated SNPs, rs1432295 and rs13034020, are both located in the same LD block (Figure 1A) [4]. For both of the SNPs the minor G allele is the risk allele. Using cross-tabulation we compared the genotype relationship between rs13034020 and rs1432295 (Figure 1B). An overlap between the genotypes of the risk and protective alleles for the two SNPs was observed for 57% of the controls and 49% of the post-therapy cHL patients.

The 2p16.1 region includes three genes, i.e. REL and the lncRNAs LINC01185 (FLJ16341) and TCONS_00002901, immediately upstream and downstream of REL, respectively. LINC01185 was expressed at very low levels according to our microarray data [15] and therefore not analyzed in this study. Expression levels of REL and TCONS_00002901 on SNP-allele based stratified LCL groups did not reveal any significant associations (Figure 1C and D). The expression levels of REL and TCONS_00002901 showed a positive correlation in LCLs of controls (R^2=0.74) and post-therapy cHL patients (R^2=0.25) (Figure 1E).

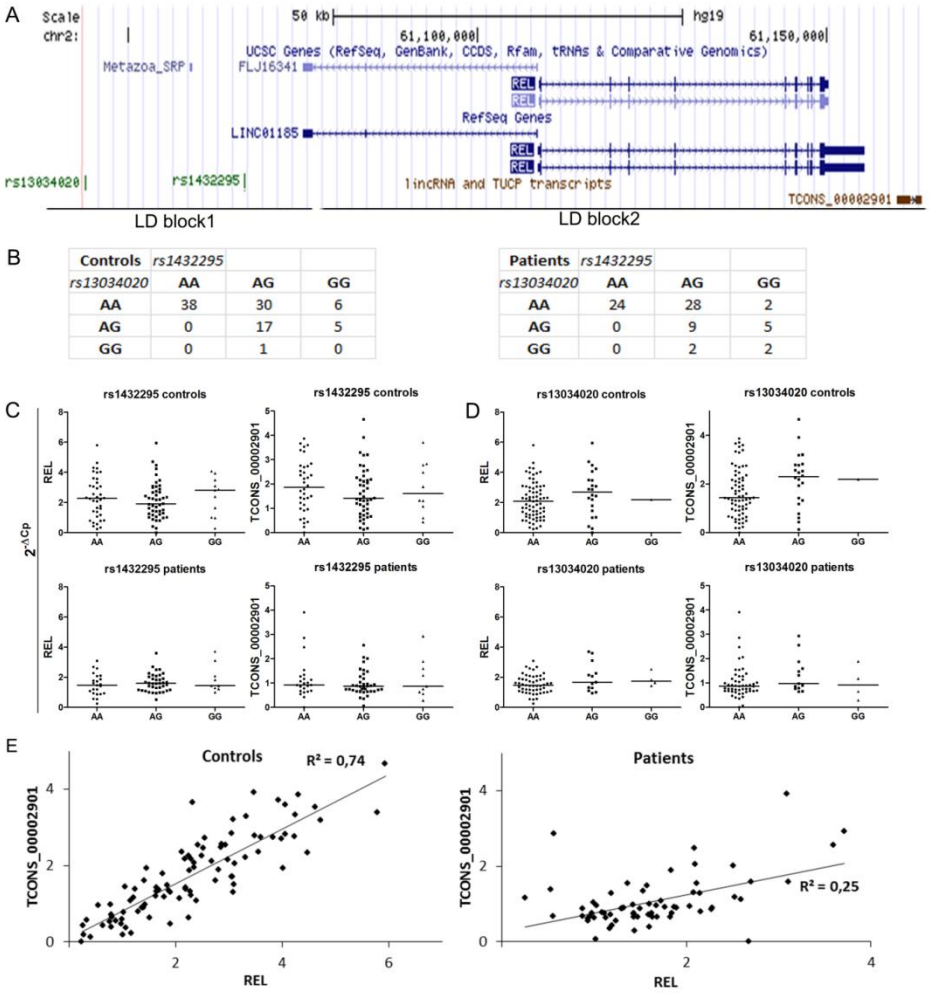
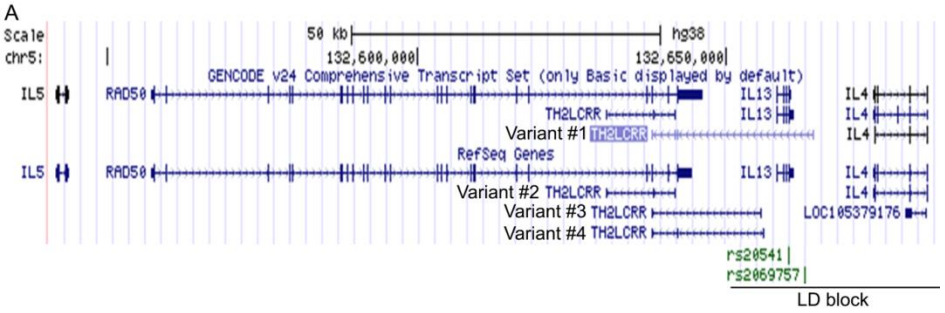


Figure 1. Schematic overview of the 2p16.1 region and expression analysis of REL and TCONS_00002901 in SNP genotype stratified LCL groups. (A) UCSC Genome browser view of the 137 kb region which contains two LD blocks [4], REL and 2 lncRNAs. The two associated SNPs are shown in green. **(B)** Cross-tabulation of the rs13034020 and rs1432295 genotype distributions for controls and post-therapy cHL patients. **(C)** In the LCLs of healthy controls and post-therapy cHL patients the REL and TCONS_00002901 expression levels were not different in the genotype stratified groups of rs13034020 and **(D)** rs1432295. The horizontal lines represent the median expression values. **(E)** Correlation plots depicting the relation between REL and TCONS_00002901 expression levels in LCLs of controls and post-therapy patients.

Expression levels of 5q31 genes in SNP genotype stratified LCLs

One of the two associated SNPs at 5q31, i.e. rs20541, is located in exon 4 of the *IL13* gene. The 2nd SNP in this locus, rs2069757, is located downstream of the 3' end of the *IL13* gene (Figure 2A). Both SNPs map within the same LD block. Cross-tabulation of the rs20541 and rs2069757 genotypes revealed consistent genotypes for the risk and protective alleles for 78% of the controls and for 76% of the post-therapy cHL patients (Figure 2B). In addition to the *IL13* gene, this locus also contains the *TH2LCRR* (T helper type 2 locus control region associated RNA) lncRNA on the opposite strand of *IL13*. This lncRNA has 4 known transcript variants. Both associated SNPs, map within the first intron of transcript variant #1 of TH2LCRR. IL13 was not expressed in LCLs (data not shown). Expression analysis of all transcript variants of TH2LCRR in the LCL groups revealed expression of variant #2 and not of the other transcript variants. No significant differences in TH2LCRR variant #2 expression levels were observed in SNP genotype stratified groups for both SNPs (Figure 2C and D).



B

Controls	*rs20541*		
rs2069757	GG	GA	AA
GG	61	17	1
GA	1	14	1
AA	0	0	1

Patients	*rs20541*		
rs2069757	GG	GA	AA
GG	42	16	1
GA	0	11	0
AA	0	0	1

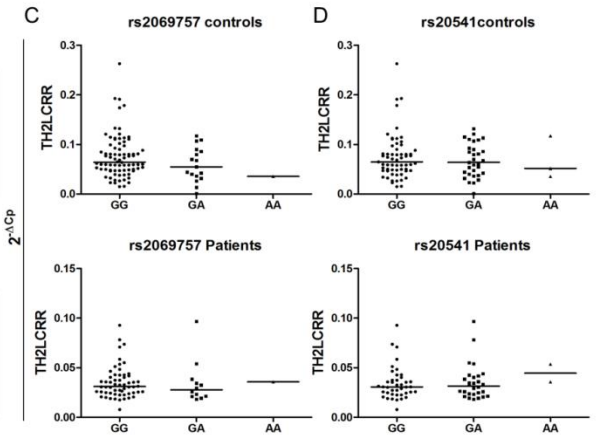
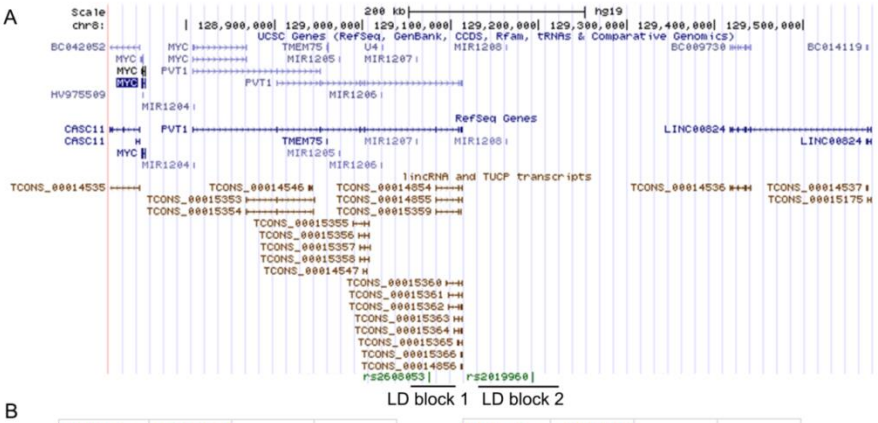


Figure 2. Schematic overview of the 5q31 region and expression analysis of the TH2LCRR lncRNA in SNP genotype stratified LCL groups. (A) UCSC Genome browser view of the 5q31 genomic region which includes IL-13 and TH2LCRR. SNPs rs2069757 and rs20541 which are both associated with cHL susceptibility are shown in green. PCR primers were designed to detect each of the four TH2-LCRR lncRNA variants separately by RT-PCR. Only variant #2 was detectable in LCLs using an exon 2 and exon 3 spanning primer set. (**B**) Cross-tabulation of the rs2069757 and rs20541 genotypes in the control (top) and post-therapy cHL patients (bottom) groups. (**C**) Analysis of the expression levels of lncRNA TH2LCRR in rs2069757 and (**D**) rs20541 genotype stratified LCL groups revealed no significant associations.

Expression levels of 8q24.21 genes in SNP genotype stratified LCLs

One of the two 8q24.21 associated SNPs, rs2608053, maps to intron 5 of *PVT1* within a 56kb LD region. The second SNP, rs2019960, maps downstream of *PVT1* within a 82kb LD region. The cross-tabulation data showed no relationship between the SNP genotypes, with consistent risk or protective allele genotypes only in 33% of the controls and cHL patients. Next to several small ncRNAs, the genomic region contains one protein-coding gene, i.e. *MYC*, and three lncRNAs, i.e. *PVT1*, *CASC11* and *LINC00824*. MYC and PVT1 were expressed in LCLs, while CASC11 and LINC00824 were not. MYC and PVT1 expression levels were not associated with either the rs2019960 or the rs2608053 genotypes (Figure 3C and D). Expression levels of MYC and PVT1 were not correlated in control (R^2=0.06) and post-therapy cHL patients derived (R^2 =0.06) LCLs (Figure 3E).

A

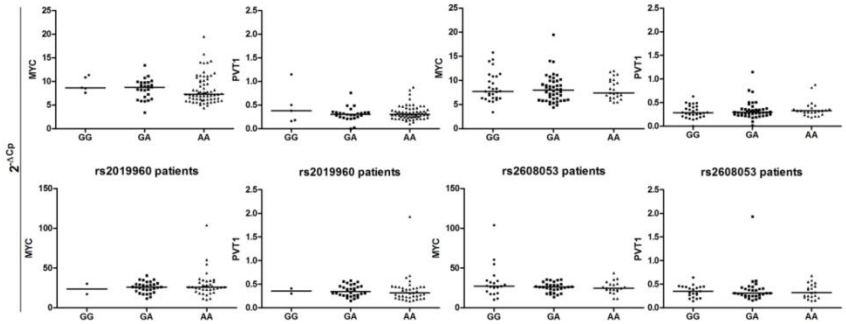


B

Controls	*rs2019960*		
rs2608053	GG	GA	AA
GG	2	6	20
GA	3	12	30
AA	0	7	17

Patients	*rs2019960*		
rs2608053	GG	GA	AA
GG	1	6	14
GA	0	14	16
AA	2	10	9

C



D

E

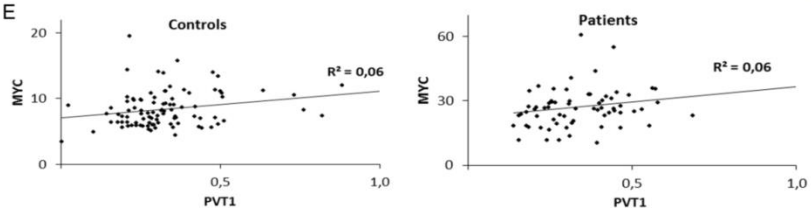
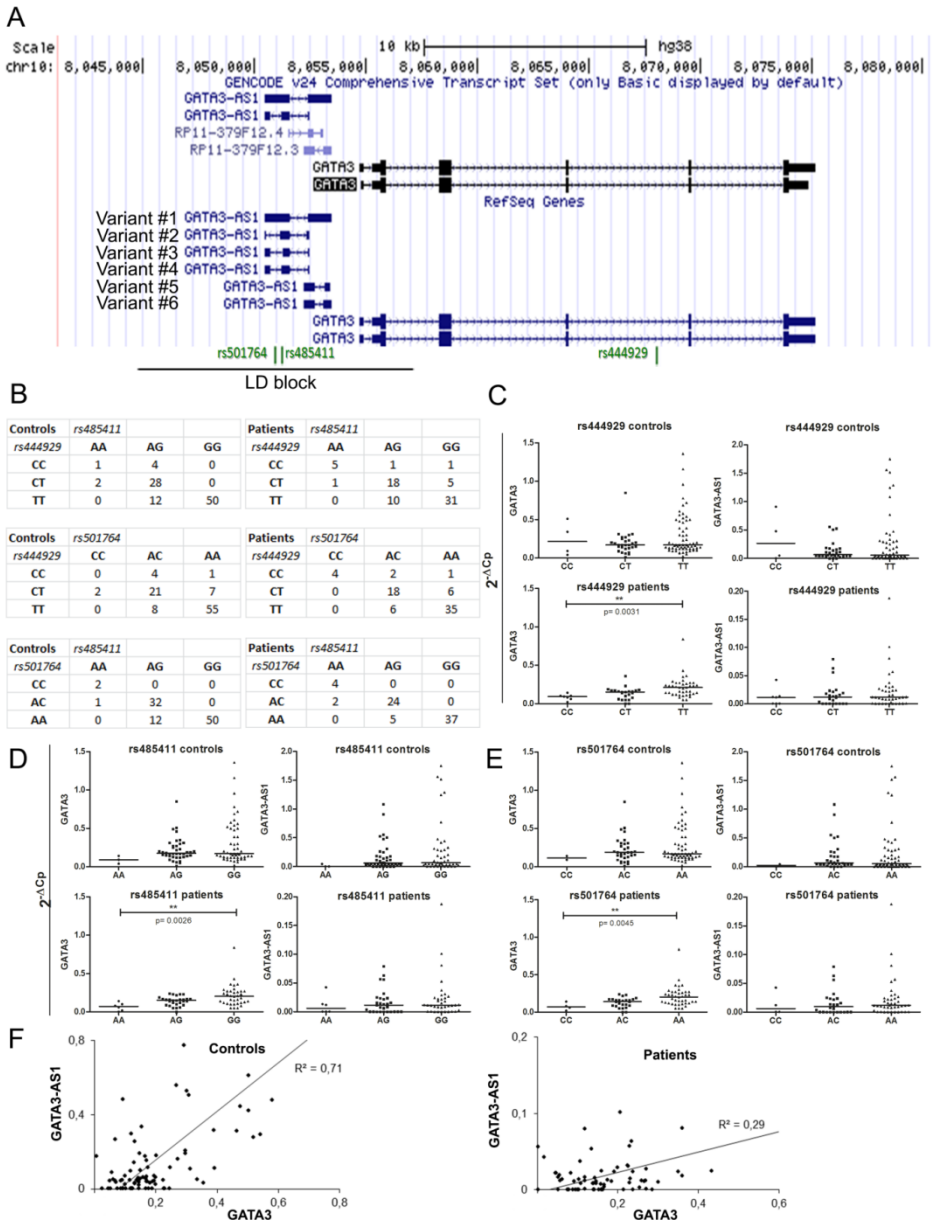


Figure 3. Schematic overview of the 8q24.21 region and expression analysis of MYC and PVT1 in SNP genotype stratified LCL groups. (**A**) UCSC Genome browser view of the 8q24.21 genomic region which includes MYC and lncRNAs CASC11, PVT1 and LINC00824. cHL susceptibility associated SNPs rs2608053 and rs2019960 are shown in green. Rs2608053 maps within a region of 56kb in LD (LD block1) and rs2019960 maps to a region of 82kb in LD (LD block2). (**B**) Cross-tabulation of the rs2019960 and rs2608053 genotype distributions. (**C**) Relationship between the expression levels of MYC, PVT1 and the genotypes of rs2019960 and (**D**) rs2608053. (**E**) Correlation plots depicting the correlation between MYC and PVT1 expression in LCLs (1 sample of the control LCL group and 2 samples of the post-therapy patient LCL group were outliers and are outside of the axis limits).

Expression levels of 10p14 genes in SNP genotype stratified LCLs

Rs501764 and rs485411 map to a 40kb LD region encompassing part of the GATA3 gene (Figure 4A). This region also contains the GATA3-AS1 lncRNA. Rs444929 maps to intron 4 of GATA3 while rs501764 and rs485411 map within 4 of the 6 transcript variants of GATA3-AS1. Rs501764 maps in exon 2 of GATA3-AS1 variant #1 and in intron 2 of the three other GATA3-AS1 transcripts variants. Rs485411 maps in exon 2 of variants #1 and #2 and in intron 2 of the other 2 variants. Cross-tabulation analysis showed that there is a strong association between the genotypes of the risk and protective alleles for all comparisons between the three SNPs in both the control and post-therapy cHL patient LCL groups. The overlap in genotypes is 75-81% for rs444929 and rs485411, 78-79% for rs444929 and rs501764 and 87-90% for rs485411 and rs501764 (Figure 4B). The genotype agreement was very high even for the rs444929 SNP that maps outside the LD block. Post-therapy cHL patient derived LCLs stratified based on genotypes of either of the three associated SNPs showed significant differences in GATA3 expression levels. LCLs derived of healthy controls did not show differences in expression levels of GATA3. For GATA3-AS1, we did not observe a significant relationship between expression levels and SNP genotype in LCLs of control or post-therapy cHL patient samples (Figure 4C-D).

Comparison of the GATA3 and GATA3-AS1 expression levels showed a strong positive correlation in the healthy control LCL group (R^2 =0.7) and a weaker correlation in the post-therapy cHL patient LCL group (R^2 =0.29) (Figure 4E).



Controls	*rs485411*			Patients	*rs485411*		
rs444929	AA	AG	GG	*rs444929*	AA	AG	GG
CC	1	4	0	CC	5	1	1
CT	2	28	0	CT	1	18	5
TT	0	12	50	TT	0	10	31

Controls	*rs501764*			Patients	*rs501764*		
rs444929	CC	AC	AA	*rs444929*	CC	AC	AA
CC	0	4	1	CC	4	2	1
CT	2	21	7	CT	0	18	6
TT	0	8	55	TT	0	6	35

Controls	*rs485411*			Patients	*rs485411*		
rs501764	AA	AG	GG	*rs501764*	AA	AG	GG
CC	2	0	0	CC	4	0	0
AC	1	32	0	AC	2	24	0
AA	0	12	50	AA	0	5	37

Figure 4. Schematic overview of the 10p14 region and expression analysis

of GATA3 and GATA3-AS1 in SNP genotype stratified LCL groups. (A) UCSC Genome browser view of the 10p14 genomic region which includes GATA3 and GATA3-AS1. The associated SNPs rs501764 and rs485411 map to a 40kb region in LD while rs444929 is located outside of this LD region. SNPs are shown in green. **(B)** Cross-tabulation of the rs444929, rs485411 and rs501764 genotype distributions. **(C)** Relationship between the expression levels of GATA3, GATA3-AS1 and the rs444929, **(D)** rs485411 and **(E)** rs501764 genotypes. **(F)** Correlation plots depicting the correlation between GATA3 and GATA3-AS1 expression in the LCL groups (11 samples in the control LCL group and 1 sample in the post-therapy patient LCL group were outliers and are outside of the axis limits).

DISCUSSION

In this study we explored whether SNPs associated with cHL susceptibility at 2p16.1, 5q31, 8q24.21 and 10p14 loci may have eQTL effects on nearby protein-coding or lncRNA genes. We observed a correlation between genotype and GATA3 gene expression levels for rs485411, rs444929 and rs501764 in LCLs of post-therapy HL patients. We did not identify eQTL effects of GATA3-AS1 or any of the other protein-coding or lncRNA genes at 2p16.1, 5q31 and 8q24.21. This suggests that the susceptibility mechanisms related to these SNP alleles do not involve eQTL effects for these genes.

However, we might have failed to detect eQTL effects, due to insufficient numbers of LCLs included in our analyses. Moreover, eQTL effects have been observed for genes that are located up to 1 MB away [16]. So, we cannot exclude the possibility that the gene relevant for the association is located further up- or downstream of the SNPs. In addition, expression of additional small ncRNAs not included in the present study and of which several are present at the 8q24.21 locus, might be associated with the SNP genotypes. Lack of expression of IL13 and some of the lncRNAs in LCLs might also implicate that we did not analyze the relevant cell type. Susceptibility effects of genetic variants associated with cHL might also be related to gene expression changes in any of the other cell types present in the microenvironment [17]. The most prominent

cell types in the microenvironment are the T cells and multiple studies show that the cross-talk between tumor cells or their precursors and T cells is crucial for the development of cHL [17-19]. Th1 cells are implicated in tumor eradication, whereas Th2 cells support the growth of HRS cells. Thus, a reduced Th1 and/or an increased Th2 polarization might alter risk to develop cHL by influencing survival of precursor HRS cells. Two of the four SNP loci harbor genes relevant for Th2 cells, i.e. GATA3 and IL13. Thus, it would be worthwhile to also perform eQTL analysis in SNP allele stratified T-cell samples.

GATA3 expression is associated with rs485411, rs501764 and rs444929 genotypes in LCLs derived from PBMCs of post-therapy cHL patients, but not in LCLs derived from PBMCs of controls. GATA-3 is a transcription factor that drives differentiation of Th0 cells into Th2 cells and plays an important role in the maintenance of a Th2 cell phenotype [20]. We observed lower levels of GATA3 in the risk allele group, which is inconsistent with the proposed beneficial effects of Th2 polarization on cHL development. GATA3 and GATA3-AS1 are co-expressed in Th2 cells [21, 22]. The correlation between GATA3 and GATA3-AS1 expression as observed in this study and in our study may reflect a *cis*-acting regulatory effect of GATA3-AS1 on GATA3. Allele-specific *cis*-effects at 10p14 provide an attractive mechanism for the association with cHL risk and deserve further investigation. The differences we observed for correlation between genotype and GATA3 gene expression levels in LCLs of post-therapy cHL patients but not in the control group, suggests that other factors influence the eQTL effects.

An eQTL effect of 10p14 (rs485411) has been observed in non-small cell lung cancer (NSCLC) [23] for *ITIH2* (inter-alpha-trypsin inhibitor heavy chain 2) [23]. *ITIH2* is located ~300 kb upstream of GATA3. ITIH2 is strongly downregulated in a variety of human solid tumors [24]. eQTL effect of rs485411 on *ITIH2* and a possible connection with abnormal immune response in cHL, deserves further investigation.

The Th2 cytokine IL-13 is important for normal B-cell survival and proliferation [25]. HRS cells frequently express IL-13 as well as the Th2 cells in their nearby environment and IL-13 has been implicated as an important growth factor for HRS cells [26]. The IL-13 missense SNP (rs20541) results in a change of arginine (R110) to glutamine (Q110) [8] and makes the protein more active than the wild type IL-13 protein [8]. A second SNP in the region (*IL13*-1112C>T, rs1800925) has been reported to increase *IL13* promoter activity in Th2 lymphocytes in allergy and asthma [27]. Thus, the susceptibility effect of the 5q31 SNPs might be related to either an increase in the activity or the levels of IL-13 produced by Th2 cells which may enhance the survival of HRS precursor cells. TH2LCRR is transcribed from a region that has previously been described in mice as Th2-locus control region (*TH2-LCR*), which is required for expression of genes encoding Th2 cytokines [28]. A more recent study [21] in human T-cell cultures demonstrated *TH2LCRR* expression specifically during Th2 polarizing conditions but not during Th1 and Th17 polarizing conditions. They showed that all *TH2LCRR* variants are Th2 lineage specific and co-expressed with *IL4*, *IL5* and *IL13*. Moreover, they demonstrated that depletion of TH2LCRR impairs the expression of Th2 cytokines, IL-4, IL-5 and IL-13 by loss of the activating histone mark H3K4Me [21]. In our study, *IL13* mRNA levels were below the detection limit while *TH2LCRR* was expressed in the LCLs, indicating that different mechanisms are employed to regulate IL13 expression. Further studies in allele stratified Th2 cells are required to determine to what extent TH2LCRR and/or IL13 expression correlates with genotypes.

For two of the SNPs a susceptibility mechanism related to B-cell survival might be most likely, i.e. the REL and MYC loci. REL is a member of the NF-κB family of transcription factors and is one of the subunits in the canonical NF-κB pathway. Several studies showed chromosomal rearrangements of the short arm of chromosome 2 encompassing the *REL* locus [29, 30]. Moreover, constitutive activation of the NF-κB pathway is a hallmark of cHL. Although we did not observe a significant relationship

between SNP genotype and expression of REL and TCONS_00002901 in LCLs, this does not preclude an effect of these SNP alleles in HRS precursor cells via other mechanisms. We did find a positive relation between expression levels of TCONS_00002901 and REL, which might be indicative of a possible cis-regulation of these genes. We did not observe eQTL effects for *PVT1* or *MYC*. Besides the known oncogenic effects of MYC there is also substantial evidence that PVT1 plays an important oncogenic role in cancer. *PVT1* is consistently co-gained with *MYC* and increased PVT1 expression is needed to maintain high MYC protein levels [31]. Murine T-cell lymphomas induced by retroviral integrations into the *PVT1* locus were shown to be driven by PVT1 overexpression which was shown to be required for maintaining high MYC levels [12]. The *PVT1* locus also contains several microRNAs (miR-1204-1208). Overexpression of miR-1204 in mouse pre-B cells enhanced MYC expression [32]. It might be interesting to determine the possible eQTL effects of the SNPs on the expression levels of these miRNAs as well.

We investigated whether cHL associated SNP alleles have eQTL effects on nearby genes, focusing on both protein-coding and non-coding transcripts in LCLs. We observed no significant relationship between expression levels and SNP genotypes at the 2p16.1, 5q31 and 8q24.21 regions in LCLs. We did identify a significant eQTL effect for all three SNPs with GATA3 expression in LCLs of post-therapy cHL patients. Alternative ways by which genetic variations may confer risk to develop cHL as well as eQTL analysis in different cell types should be a focus in follow-up experiments.

Table S1. TaqMan SNP assay used in this study.

SNP ID	Assay ID	Gene locus	Genotype VIC/FAM
rs1432295	C_2080978_10	2p16.1	A/G
rs13034020	C_31842483_10	2p16.1	A/G
rs2069757	C_16176550_10	5q31	A/G
rs20541	C_2259921_20	5q31	A/G
rs2019960	C_2767164_10	8q24.21	C/T
rs2608053	C_15904124_10	8q24.21	C/T
rs444929	C_3182060_10	10p14	C/T
rs485411	C_11647753_20	10p14	C/T
rs501764	C_1575387_10	10p14	G/T

Table S2. Gene expression assay IDs and primer sets used in this study.

Transcripts	Gene Locus	Forward and reverse sequences or gene expression assay ID
REL	2p16.1	Hs00968440_m1
TCONS_00002901	2p16.1	F: 5'-GCAGAGACCATGGGAGTCAGA-3' R: 5'-CCACATGATCCAGCAATCCTACT-3'
IL13	5q31	Hs00174379_m1
TH2LCRR	5q31	F: 5'-CGCTGCTTTACTTGGGTGAGT-3' R: 5'-CGATGCAGTGCTGGACAAAA-3'
MYC	8q24.21	F: 5'- CACCAGCAGCGACTCTGA-3' R: 5'- ATCCAGACTCTGACCTTTTGC-3'
PVT1	8q24.21	F: 5'-ACTCAGCAGCAAGCACCTGTT-3' R: 5'-GGATTTTTCAAAGGCAGTTTGG-3'
GATA3	10p14	Hs00231122_m1
GATA3-AS1	10p14	F: 5'-CCACGCGTAAGACCAAAAGC-3' R: 5'-GGACTTCTCTCTTCGGCTCTGA-3'
TBP	6q27	F: 5'- GCCCGAAACGCCGAATAT-3' R: 5'- CCGTGGTTCGTGGCTCTCT-3'

REFERENCES

1. Kerzin-Storrar L, Faed MJ, MacGillivray JB, Smith PG: Incidence of familial Hodgkin's disease. Br J Cancer 1983, 47:707-712.

2. Mack TM, Cozen W, Shibata DK, Weiss LM, Nathwani BN, Hernandez AM, Taylor CR, Hamilton AS, Deapen DM, Rappaport EB: Concordance for Hodgkin's disease in identical twins suggesting genetic susceptibility to the young-adult form of the disease. N Engl J Med 1995, 332:413-418.

3. Urayama KY, Jarrett RF, Hjalgrim H, Diepstra A, Kamatani Y, Chabrier A, Gaborieau V, Boland A, Nieters A, Becker N, Foretova L, Benavente Y, Maynadie M, Staines A, Shield L, Lake A, Montgomery D, Taylor M, Smedby KE, Amini RM, Adami HO, Glimelius B, Feenstra B, Nolte IM, Visser L, van Imhoff GW, Lightfoot T, Cocco P, Kiemeney L, Vermeulen SH, Holcatova I, Vatten L, Macfarlane GJ, Thomson P, Conway DI, Benhamou S, Agudo A, Healy CM, Overvad K, Tjonneland A, Melin B, Canzian F, Khaw KT, Travis RC, Peeters PH, Gonzalez CA, Quiros JR, Sanchez MJ, Huerta JM, Ardanaz E, Dorronsoro M, Clavel-Chapelon F, Bueno-de-Mesquita HB, Riboli E, Roman E, Boffetta P, de Sanjose S, Zelenika D, Melbye M, van den Berg A, Lathrop M, Brennan P, McKay JD: Genome-wide association study of classical Hodgkin lymphoma and Epstein-Barr virus status-defined subgroups. J Natl Cancer Inst 2012, 104:240-253.

4. Enciso-Mora V, Broderick P, Ma Y, Jarrett RF, Hjalgrim H, Hemminki K, van den Berg A, Olver B, Lloyd A, Dobbins SE, Lightfoot T, van Leeuwen FE, Forsti A, Diepstra A, Broeks A, Vijayakrishnan J, Shield L, Lake A, Montgomery D, Roman E, Engert A, von Strandmann EP, Reiners KS, Nolte IM, Smedby KE, Adami HO, Russell NS, Glimelius B, Hamilton-Dutoit S, de Bruin M, Ryder LP, Molin D, Sorensen KM, Chang ET, Taylor M, Cooke R, Hofstra R, Westers H, van Wezel T, van Eijk R, Ashworth A, Rostgaard K, Melbye M, Swerdlow AJ, Houlston RS: A genome-wide association study of Hodgkin's lymphoma identifies new susceptibility loci at 2p16.1 (REL), 8q24.21 and 10p14 (GATA3). Nat Genet 2010, 42:1126-1130.

5. Cozen W, Li D, Best T, Van Den Berg DJ, Gourraud PA, Cortessis VK, Skol AD, Mack TM, Glaser SL, Weiss LM, Nathwani BN, Bhatia S, Schumacher FR, Edlund CK, Hwang AE, Slager SL, Fredericksen ZS, Strong LC, Habermann TM, Link BK, Cerhan JR, Robison LL, Conti DV, Onel K: A genome-wide meta-analysis of nodular sclerosing Hodgkin lymphoma identifies risk loci at 6p21.32. Blood 2012, 119:469-475.

6. Frampton M, da Silva Filho MI, Broderick P, Thomsen H, Forsti A, Vijayakrishnan J, Cooke R, Enciso-Mora V, Hoffmann P, Nothen MM, Lloyd A, Holroyd A, Eisele L, Jockel KH, Ponader S, von Strandmann EP, Lightfoot T, Roman E, Lake A, Montgomery D, Jarrett RF, Swerdlow AJ, Engert A, Hemminki K, Houlston RS: Variation at 3p24.1 and 6q23.3 influences the risk of Hodgkin's lymphoma. Nat Commun 2013, 4:2549.

7. Cozen W, Timofeeva MN, Li D, Diepstra A, Hazelett D, Delahaye-Sourdeix M, Edlund CK, Franke L, Rostgaard K, Van Den Berg DJ, Cortessis VK, Smedby KE, Glaser SL, Westra HJ, Robison LL, Mack TM, Ghesquieres H, Hwang AE, Nieters A, de Sanjose S, Lightfoot T, Becker N, Maynadie M, Foretova L, Roman E, Benavente Y, Rand KA, Nathwani BN, Glimelius B, Staines A, Boffetta P, Link BK, Kiemeney L, Ansell SM, Bhatia S, Strong LC, Galan P, Vatten L, Habermann TM, Duell EJ, Lake A, Veenstra RN, Visser L, Liu Y, Urayama KY, Montgomery D, Gaborieau V, Weiss LM, Byrnes G, Lathrop M, Cocco P, Best T, Skol AD, Adami HO, Melbye M, Cerhan JR, Gallagher A, Taylor GM, Slager SL, Brennan P, Coetzee GA, Conti DV, Onel K, Jarrett RF, Hjalgrim H, van den Berg A, McKay JD: A meta-analysis of Hodgkin lymphoma reveals 19p13.3 TCF3 as a novel susceptibility locus. Nat Commun 2014, 5:3856.

8. Vladich FD, Brazille SM, Stern D, Peck ML, Ghittoni R, Vercelli D: IL-13 R130Q, a common variant associated with allergy and asthma, enhances effector mechanisms essential for human allergic inflammation. J Clin Invest 2005, 115:747-754.

9. Hindorff LA, Sethupathy P, Junkins HA, Ramos EM, Mehta JP, Collins FS, Manolio TA: Potential etiologic and functional implications of genome-wide association loci for human diseases and traits. Proc Natl Acad Sci U S A 2009, 106:9362-9367.

10. Hrdlickova B, de Almeida RC, Borek Z, Withoff S: Genetic variation in the non-coding genome: Involvement of micro-RNAs and long non-coding RNAs in disease. Biochim Biophys Acta 2014, 1842:1910-1922.

11. Kumar V, Westra HJ, Karjalainen J, Zhernakova DV, Esko T, Hrdlickova B, Almeida R, Zhernakova A, Reinmaa E, Vosa U, Hofker MH, Fehrmann RS, Fu J, Withoff S, Metspalu A, Franke L, Wijmenga C: Human disease-associated genetic variation impacts large intergenic non-coding RNA expression. PLoS Genet 2013, 9:e1003201.

12. Graham M, Adams JM, Cory S: Murine T lymphomas with retroviral inserts in the chromosomal 15 locus for plasmacytoma variant translocations. Nature 1985, 314:740-743.

13. Hui-Yuen J, McAllister S, Koganti S, Hill E, Bhaduri-McIntosh S: Establishment of Epstein-Barr Virus Growth-transformed Lymphoblastoid Cell Lines. J Vis Exp 2011, (57):3321. doi:10.3791/3321.

14. Winkle M, van den Berg A, Tayari M, Sietzema J, Terpstra M, Kortman G, de Jong D, Visser L, Diepstra A, Kok K, Kluiver J: Long noncoding RNAs as a novel component of the Myc transcriptional network. FASEB J 2015.

15. Tayari M, Winkle M, Kortman G, Sietzema J, de Jong D, Terpstra M, Mestdagh P, Kroese FG, Visser L, Diepstra A, Kok K, van den Berg A, Kluiver J: Long Noncoding RNA Expression Profiling in Normal B-Cell Subsets and Hodgkin Lymphoma Reveals Hodgkin and Reed-Sternberg Cell-Specific Long Noncoding RNAs. Am J Pathol 2016.

16. Nica AC, Dermitzakis ET: Expression quantitative trait loci: present and future. Philos Trans R Soc Lond B Biol Sci 2013, 368:20120362.

17. Scott DW, Steidl C: The classical Hodgkin lymphoma tumor microenvironment: macrophages and gene expression-based modeling. Hematology Am Soc Hematol Educ Program 2014, 2014:144-150.

18. Steidl C, Connors JM, Gascoyne RD: Molecular pathogenesis of Hodgkin's lymphoma: increasing evidence of the importance of the microenvironment. J Clin Oncol 2011, 29:1812-1826.

19. Aldinucci D, Celegato M, Casagrande N: Microenvironmental interactions in classical Hodgkin lymphoma and their role in promoting tumor growth, immune escape and drug resistance. Cancer Lett 2016, 380:243-252.

20. Ho IC, Tai TS, Pai SY: GATA3 and the T-cell lineage: essential functions before and after T-helper-2-cell differentiation. Nat Rev Immunol 2009, 9:125-135.

21. Spurlock CF,3rd, Tossberg JT, Guo Y, Collier SP, Crooke PS,3rd, Aune TM: Expression and functions of long noncoding RNAs during human T helper cell differentiation. Nat Commun 2015, 6:6932.

22. Hu G, Tang Q, Sharma S, Yu F, Escobar TM, Muljo SA, Zhu J, Zhao K: Expression and regulation of intergenic long noncoding RNAs during T cell development and differentiation. Nat Immunol 2013, 14:1190-1198.

23. Wu X, Wang L, Ye Y, Aakre JA, Pu X, Chang GC, Yang PC, Roth JA, Marks RS, Lippman SM, Chang JY, Lu C, Deschamps C, Su WC, Wang WC, Huang MS, Chang DW, Li Y, Pankratz VS, Minna JD, Hong WK, Hildebrandt MAT, Hsiung CA, Yang P: Genome-wide association study of genetic predictors of overall survival for non-small cell lung cancer in never smokers. Cancer Res 2013, 73:4028-4038.

24. Hamm A, Veeck J, Bektas N, Wild PJ, Hartmann A, Heindrichs U, Kristiansen G, Werbowetski-Ogilvie T, Del Maestro R, Knuechel R, Dahl E: Frequent expression loss of Inter-alpha-trypsin inhibitor heavy chain (ITIH) genes in multiple human solid tumors: a systematic expression analysis. BMC Cancer 2008, 8:25-2407-8-25.

25. Ford D, Sheehan C, Girasole C, Priester R, Kouttab N, Tigges J, King TC, Luciani A, Morgan JW, Maizel AL: The human B cell response to IL-13 is dependent on cellular phenotype as well as mode of activation. J Immunol 1999, 163:3185-3193.

26. Skinnider BF, Elia AJ, Gascoyne RD, Trumper LH, von Bonin F, Kapp U, Patterson B, Snow BE, Mak TW: Interleukin 13 and interleukin 13 receptor are frequently expressed by Hodgkin and Reed-Sternberg cells of Hodgkin lymphoma. Blood 2001, 97:250-255.

27. Cameron L, Webster RB, Strempel JM, Kiesler P, Kabesch M, Ramachandran H, Yu L, Stern DA, Graves PE, Lohman IC, Wright AL, Halonen M, Klimecki WT, Vercelli D: Th2 cell-selective enhancement of human IL13 transcription by IL13-1112C>T, a polymorphism associated with allergic inflammation. J Immunol 2006, 177:8633-8642.

28. Lee GR, Spilianakis CG, Flavell RA: Hypersensitive site 7 of the TH2 locus control region is essential for expressing TH2 cytokine genes and for long-range intrachromosomal interactions. Nat Immunol 2005, 6:42-48.

29. Joos S, Menz CK, Wrobel G, Siebert R, Gesk S, Ohl S, Mechtersheimer G, Trumper L, Moller P, Lichter P, Barth TF: Classical Hodgkin lymphoma is characterized by recurrent copy number gains of the short arm of chromosome 2. Blood 2002, 99:1381-1387.

30. Martin-Subero JI, Gesk S, Harder L, Sonoki T, Tucker PW, Schlegelberger B, Grote W, Novo FJ, Calasanz MJ, Hansmann ML, Dyer MJ, Siebert R: Recurrent involvement of the REL and BCL11A loci in classical Hodgkin lymphoma. Blood 2002, 99:1474-1477.

31. Tseng YY, Moriarity BS, Gong W, Akiyama R, Tiwari A, Kawakami H, Ronning P, Reuland B, Guenther K, Beadnell TC, Essig J, Otto GM, O'Sullivan MG, Largaespada DA, Schwertfeger KL, Marahrens Y, Kawakami Y, Bagchi A: PVT1 dependence in cancer with MYC copy-number increase. Nature 2014, 512:82-86.

32. Huppi K, Volfovsky N, Runfola T, Jones TL, Mackiewicz M, Martin SE, Mushinski JF, Stephens R, Caplen NJ: The identification of microRNAs in a genomically unstable region of human chromosome 8q24. Mol Cancer Res 2008, 6:212-221.